Conformational analysis of protein and peptide sequence units

Ph.D. thesis

ILONA HUDÁKY

Supervisor: Prof. Dr. András Perczel

Chemistry Doctoral School
Head: Prof. Dr. György Inzelt

Program: Theoretical and Physical Chemistry, Stuctural Chemistry
Head: Prof. Dr. Péter Surján

Eötvös Loránd University
Institute of Chemistry

2009

INTRODUCTION

As a Ph.D. work, I carried out quantum chemical modelling and database analysis of peptide sequence units that are conformationally restricted due to their ring systems, such as proline, prolylproline and vicinal disulfide bridges. Biological importance of the selected model compounds is expressed in their presence in collagen and in receptors at the neuromuscular junction. My investigations are connected to the conformational studies on various amino acid residues and alanylalanine, and to the nomenclature of amino acid backbone structures previously worked out in our research group (see **Figure 1**). In searching for minima, the entire conformational space of the model peptide is explored, thus *cis* amid and versatile side-chain orientations are also respected. Such applied quantum chemical calculations can either conformationally characterize a "building block" of polypeptides, or test the appropriateness of approaches at different levels of theory, or compare the reliability of differing experimental methods.

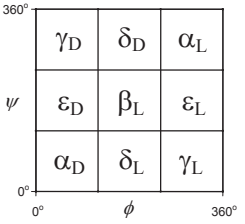

Figure 1 Labelling of amino acid residue conformers according to the systematic division of the Ramachandran diagram as worked out in our research group.

Scope

1. Elaboration of a conformational nomenclature that is able to assign diamides as well as longer protein chains, and seems simple and self-consistent in order to help the handling of structural data and consequently the cooperation between the different scientific approaches to proteins and peptides.

2. Exploring whether calculations using the PCM solvent model can deliver α-helical and poliproline II type conformers of the For-L-Ala-NH_2 model peptide, α_L and ε_L, which are not genuine minima by gas phase quantum chemical calculations.

3. Building up complete conformational libraries of the proline residue by *ab initio* calculations and database analysis.

4. Conformational studies on the prolylproline sequence unit using gas phase and PCM calculations at various *ab initio* and DFT levels and information retrieved from databases.

5. Investigating the structural restrictions strained on a cysteinylcysteine sequence unit by a vicinal disulfide bridge (see **Figure 2**).

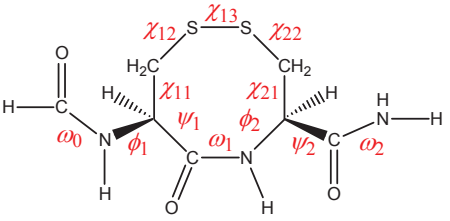

Figure 2 A model of the vicinal disulfide bridge, a rare structural motif in proteins. Model peptide HCO–*ox*-[L-Cys–L-Cys]–NH_2 and its characteristic torsional angles.

METHODS

Quantum chemical calculations were carried out using GAUSSIAN98 and GAUSSIAN03 program packages. The GDIIS algorithm was applied at structure optimizations. Solvent effects were accounted for using the Polarizable Continuum Model (PCM).

For statistical analysis of empirical information three databases were utilized: (i) the Cambridge Structural Database (CSD) for X-ray data on peptides, (ii) the Protein Data Bank (PDB) for large number of protein structures refined by NMR and X-ray, (iii) and PDB SELECT for homology filtered protein data. Geometrical and population data were compared to relevant results from quantum chemical calculations.

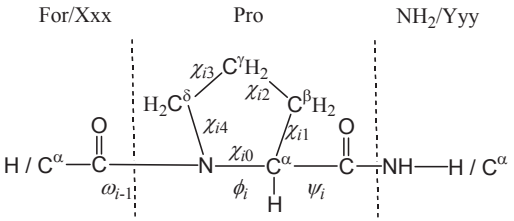

Figure 3 Selected torsional angles in the model peptide HCO–L-Pro–NH_2 and in proline residues of proteins and peptides.

The following formula of the pseudorotation was applied to the pyrrolidine ring of proline:

$$\chi_{ij} = A_i \cos(P_i + 144° j)\,,$$

where the dimension of A_i and P_i is degree, and A_i is always positive. Pseudorotational amplitude, A_i, and pseudorotational phase angle, P_i, of the i^{th} proline residue in the sequence is calculated from the five χ_{ij} torsional angles (see **Figure 3)** in such a way that for any integer j between 0 and 4 the above formula is satisfied.

RESULTS

1. Conformational nomenclature

A complete nomenclature of conformers is proposed for the general structural description of peptides of any length ranging from amino acid diamides to protein sequences. Apart from the traditionally considered backbone fold, amide *cis-trans* isomers and side-chain orientations are also involved. The code refers to the composite amino acid residues from the N-terminal to the C terminal, and applies different symbols for the characterization of amide bonds, backbone fold (ϕ_i, ψ_i pairs) and side-chains. I applied this nomenclature to proline, to prolylproline and in an extended version to vicinal disulfide conformers.

2. Alanine diamide studied using a solvent model

Ramachandran surfaces of the HCO-L-Ala-NH_2 peptide model at RHF/PCM/3-21G, RHF/PCM/6-31+G(d) and B3LYP/PCM/6-31+G(d) levels of theory are presented. Minima at the above three levels together with B3LYP/PCM/6-311++G(d,p) are optimized. Frequency calculations carried out at RHF/PCM/6-31+G(d) and B3LYP/PCM/6-31+G(d) levels of theory prove that the relevant structures are genuine minima. Results are compared to experimental data of 5393 alanine residues retrieved from PDB SELECT. PCM optimization deliver α-helical and poliproline II type conformers, α_L and ε_L. PCM optimization proves to be advantageous over single point PCM energy calculations, gas phase approaches and the application of explicit solvent molecules.

3. Proline

Analyses of peptide sequence X-ray data from CSD and *ab initio* computations at 3-21G and 6-31+G(d) levels of theory on the HCO-L-Pro-NH_2 peptide model provide complete conformational libraries of proline (see **Figure 4**). Beside *cis*-*trans* isomerism and backbone conformation, ring puckering was also analyzed quantitatively, through the calculation of pseudorotational coordinates.

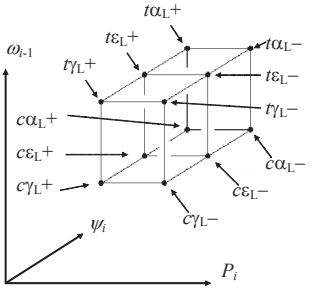

Figure 4 The twelve ideal conformers of a proline residue according to their positions in the conformational space defined by independent variables ω_{i-1}, P_i and ψ_i.

It was found that corresponding geometrical data of conformers present in both experimental and computational sets differ from each other with about the value of the standard deviation within the experimental set. Out of the twelve ideal conformers not exactly the same are present in different sets, and the population data derived from energy calculation do not reflect population data gained from the conformational analysis of synthetic model peptides retrieved from CSD. Elongation of the model peptide in such a way that could allow more types of hydrogen bonds would probably lower the expected population of *trans* γ_L conformers and thus it might give a better reflection of the conformational distribution of proline residues in peptides and proteins. Solvent models would make further improvement.

4. Prolylproline

Quantum chemical calculations applying the polarizable continuum model on HCO─L-Pro─L-Pro–NH_2 succeeded in the modelling **Pro─Pro** sequence units of polypeptides, as the same six folds are most preferred out of the 36 ideally possible backbone conformers. The orders of abundances differ, but the six favoured structures describe the great majority of prolylproline units at all applied approaches (varying from 72 to 100%).

Prolylproline either folds into elongated structures or bends into a β-turn. Although the polyproline II unit ($t\varepsilon_L t\varepsilon_L$) is dominant, other elongated folds with higher relative energy and without traditional nomenclature, like $t\varepsilon_L f\alpha_L$, $c\varepsilon_L t\varepsilon_L$ and $t\varepsilon_L f\gamma_L$ are also important in proteins. Several β-turn types are accessible for **Pro–Pro**, such as $t\varepsilon_L c\alpha_L$ (VIa), $t\alpha_L t\alpha_L$ (I or III), $t\varepsilon_L c\varepsilon_L$ (VIb), and unclassified turns of higher energy ($c\varepsilon_L c\varepsilon_L$ and $c\varepsilon_L c\alpha_L$).

Prolylproline is a sequence unit where local backbone conformational preference largely overcomes any global cooperativity. While the side chains impose certain conformational restrictions, their packing provides some sort of flexibility. This is a reason why proline residues are so frequently found to contribute to the segmental motions of disordered proteins or domains. The intrinsic probability and the found abundance of elongated structures is 87%, while that of β-turns is 13% in the **Pro–Pro** sequence units.

Tracing of the pathway of a *cis-trans* isomerization proved the proposed connections between the transition state and two relaxed structures of the model peptide For-L-Pro-L-Pro-NH_2. Backbone conformational changes, such as rotations around the ω torsional angle, are assisted by movements of the side chains, in this case pseudorotational changes of the pyrrolidine rings. The pseudorotational formula is found to fit the pyrrolidine rings not only in the minima but also in the transition state and other points of the investigated path. Throughout the pathway, the structure is stabilized by a hydrogen bond.

5. Vicinal disulfide bridges

Exhaustive conformational search at RHF/3-21G* followed by subsequent optimization at B3LYP/6-31+G(d) delivered 14 low energy conformers of the HCO–*ox*-[L-Cys–L-Cys]–NH_2 model peptide. The reliability of the results was examined by frequency calculation at B3LYP/6-31+G(d), energy calculation and optimization at B3LYP/6-311++G(d,p), as well as optimization at B3LYP/PCM/6-31+G(d). Experimental data on the ***ox*-[Cys–Cys]** sequence unit from databases and literature were successfully related to the above computed conformers. It is found that the ***ox*-[Cys–Cys]** unit is able to adopt types I, II, VIa, VIb and VIII β-turn structures, while without the vicinal disulfide bridge the **Cys–Cys** sequence unit is not constrained to form a β-turn. We believe that vicinal disulfide bridges may stabilize otherwise high energy β-turn structures, such as type VIII, as well as *cis* amide containing types VIa and VIb.

Applicability of the results

1. This nomenclature is open to alteration, *i.e.* backbone folds may be denoted according to any other convenient division of the Ramachandran map, while the system of one-character notations referring to amide bonds (***t*** or ***c***) and conformationally relevant side-chain torsional angles (+, –, a or s) are conserved.

2. The nomenclature, as applied to dipeptide backbone conformers, can identify β-turns traditionally not classified, and it can also differentiate between the numerous types of extended/elongated structures.

3. In the case of searches for peptide conformers, PCM proves a reasonable alternative to the application of explicit solvent molecules into the quantum chemically studied model system.

4. The pseudorotational formula as applied to the pyrrolidine ring of proline offers a quantitative way for the comparison of ring conformers and for the tracing of strains on the ring caused by other conformational changes.

5. Quantum chemical calculations modelling **Pro–Pro** and ***ox*-[Cys–Cys]** sequence unites give basis to critical interpretation of experimental data on proteins containing such motifs. This is of high importance in the case of proline-rich intrinsically unstructured proteins.

ACKNOWLEDGEMENT

Special thanks to my supervisor, Prof. Dr. András Perczel.

PUBLICATIONS

The thesis is based on the following publications:

(1) **Hudáky I**, Baldoni HA, Perczel A. Peptide Models XXXVIII. Proline Conformers from X-ray crystallographic database and from ab initio computations. *J Mol Struct (THEOCHEM)* 2002;582:233-249.

(2) **Hudáky I**, Perczel A. Ab initio analysis of the conformational changes of the prolyl-proline model peptide. *J Mol Struct (THEOCHEM)* 2003;630:135-140.

(3) **Hudáky I**, Kiss R, Perczel A. A nomenclature of peptide conformers. *J Mol Struct (THEOCHEM)* 2004;675:177-183.

(4) **Hudáky I**, Gáspári Z, Carugo O, Čemažar M, Pongor S, Perczel A. Vicinal disulfide bridge conformers by experimental methods and by ab initio and DFT molecular computations. *Proteins-Structure Function and Bioinformatics* 2004;55(1):152-168.

(5) **Hudáky I**, Hudáky P, Perczel A. Solvation model induced structural changes in peptides. A quantum chemical study on Ramachandran surfaces and conformers of alanine diamide using the polarizable continuum model. *J Comp Chem* 2004;25(12):1522-1531.

(6) **Hudáky I**, Perczel A. Prolylproline unit in model peptides and in fragments from databases. *Proteins-Structure Function and Bioinformatics* 2008;70:1389-1407.

Further related publications:

(7) **Hudáky I**, Perczel A. Ab initio conformational analysis of the prolyl-proline peptides and their potential to modify the main-chain fold of proteins. In *Peptides 2000 (Proceedings of the Twenty-Sixth European Peptide Symposium)*, ed. by Martinez J and Fehrentz JA, EDK, Paris, France, 2001;417-418

(8) **Hudáky I,** Perczel A. Ab initio conformational analysis of the cis-trans isomerization of the prolyl-proline peptide. In *Peptides 2002 (Proceedings of the Twenty-Seventh European Peptide Symposium)*, ed. by Benedetti E and Pedone C, Edizioni Ziino, Napoli, Italy, 2003;850-851.

(9) Carugo O, Čemažar M, Zahariev S, **Hudáky I**, Gáspári Z, Perczel A, Pongor S. Vicinal disulfide turns. *Protein Engineering* 2003;16(9):637-639.

(10) Gáspári Z, **Hudáky I**, Czajlik A, Perczel A. Is there an excuse for the non-conformist? Notes on the calculated energies, atom-atom contacts and natural abundance of the different conformers of alanine in proteins. *J Mol Struct (THEOCHEM)* 2004;675:141-148.

Publications not related to the thesis:

(11) Hudáky P; **Hudáky I**; Perczel A. Peptide models XXXV. Protonated and deprotonated N-Formyl-L-histidinamide: an ab initio study on side-chain potential energy surfaces of all major backbone conformers. *J Mol Struct (THEOCHEM)* 2002;583:199-213.

(12) Bágyi I; Balogh B; Czajlik A; Élias O; Gáspári Z; Gergely V; **Hudáky I**; Hudáky P; Kalászi A; Károlyházy L; Keserű K; Kiss R; Krajsovszky G; Láng B; Nagy T; Rácz A; Szentesi A; Tábi T; Tapolcsányi P; Vaik J; Koo JCP; Chass GA; Farkas Ö; Perczel A; Mátyus P. Generation and analysis of the conformational potential energy surfaces of N-acetyl-N-methyl-L-alanine-N'-methylamide. An exploratory ab initio study. *J Mol Struct (THEOCHEM)* 2003;625:121-136.

(13) Lam JSW, Koo JCP, **Hudáky I**, Varro A, Papp JG, Penke B, Csizmadia IG. Predicting the conformational preferences of N-acetyl-4-hydroxy-L-proline-N'-methylamide from the proline residue. *J Mol Struct (THEOCHEM)* 2003;666-667:285-289.